Interfaces with Other Disciplines

Determinants of mutual fund underperformance: A Bayesian stochastic frontier approach

Jan Annaert a,b,*, Julien van den Broeck b,1, Rudi Vander Vennet a,2

a Ghent University, W. Wilsonplein 5 D, 9000 Ghent, Belgium
b University of Antwerp, Prinsstraat 13, 2000 Antwerp, Belgium

Received 11 December 2000; accepted 25 June 2002

Abstract

The purpose of this paper is to identify ex ante fund statistics that can be related to future performance of European equity funds. In an efficient market setting, actively managed portfolios cannot outperform a passive benchmark strategy. However, purely by chance, some funds outperform their benchmark ex post, making the identification of performance determinants a difficult task. To alleviate this problem, we decompose the return deviation from its expected return into a noise component and an efficiency term, which is 100% if the fund exhibits no underperformance. The decomposition is based on the Bayesian frontier approach. We find evidence that fund efficiency is positively related to fund size and historical performance, the latter being solely due to the poorly performing funds. We fail to find a link between fund age and performance.

© 2002 Elsevier B.V. All rights reserved.

Keywords: Finance; Investment analysis; Mutual funds; Stochastic frontier

1. Introduction

Institutional portfolios manage an ever-increasing part of investors’ savings. Although they offer small investors the opportunity to invest in diversified portfolios and free them to a large extent from the burden to make allocation decisions, it is important to know at what cost these alleged advantages are offered. Performance measurement tries to tackle this issue: Are funds which are actively managed by professional managers able to achieve higher returns than passively managed funds, or do the former merely incur additional transaction costs, thus lowering return? Related to this topic is the question how investors should organise the allocation of their savings across different mutual funds: Is it possible to identify managers who consistently beat either their benchmarks or their peers? If a reliable relation exists between a fund’s performance and some of its observable characteristics, this information should be valuable to investors to guide their investment allocation across funds.
The mutual fund industry has been a topic of intensive research in recent years. The literature has broadly come to a consensus that mutual funds, on average, cannot beat the market (or passively managed portfolios), an observation that clearly aligns with the efficient market hypothesis (EMH) (Fama, 1970, 1991). Nevertheless, there are indications that some ranking of the future investment performance of managers can be achieved using predetermined variables. Performance—both measured in an absolute way and on a risk-adjusted basis—is related to past performance and fund characteristics such as expenses and size (see the discussion and references in Section 3). Nevertheless, in a simulation study, Kothari and Warner (2001) show that the sampling error of standard performance measures is large, also when no abnormal performance is present. Therefore, it is not unlikely that, due to sampling noise, some funds will be found to outperform their benchmarks, even when their managers have no real ability to outperform. This may obscure the relation between performance and fund characteristics.

To alleviate this problem, a different approach is followed in this paper. The well-documented result that mutual funds as a whole cannot outperform passive benchmarks is treated as an assumption supported by theory. In this setting, funds can only aspire to perform as well as an efficient combination of passive portfolios. In this case, investors do not hold mutual funds to achieve higher performance but, e.g., to relieve them from the burden of managing a well-diversified portfolio themselves or to obtain some potential tax advantages. We say that such mutual funds are 100% efficient or that they plot on the efficient frontier. If the fund generates too many expenses, if the fund sponsor demands too high a fee, or if wrong investment decisions have been taken, the fund will deviate from this efficient frontier. These funds will be said to be less than 100% efficient. In this paper, the efficient frontier as well as the efficiency measure is considered to be of a stochastic nature. This approach has been labelled the 'stochastic frontier method' (Meeusen and van den Broeck, 1977; Aigner et al., 1977). The methodology assumes a composed error framework, which attaches a normal distribution to the frontier to model stochastic noise and a one-sided distribution (e.g., a gamma distribution) to the efficiency measure. The estimation of the latter will result in a non-contaminated efficiency term, which can be related to particular fund characteristics in order to determine the sources of mutual fund performance, which is the aim of this paper.

Our analysis reveals that size and historical performance are related to fund efficiency. Larger funds tend to exhibit a higher degree of efficiency than small funds. This finding suggests the presence of economies of scale, but it may also be related to relatively larger capital inflows into successful funds. We also find that poor performers tend to be less efficient in a subsequent period. However, relatively good performance does not necessarily imply higher levels of efficiency. Finally, we do not find cross-sectional efficiency differences across age groups.

The remainder of this paper is organised as follows. In the next two sections we discuss the literature on performance evaluation and the relation between performance and fund characteristics. We measure performance using Jensen’s alpha from a standard market model. Section 4 describes the data set. The stochastic frontier methodology is described in Section 5 and the estimation results are presented in Section 6. Section 7 relates the estimated efficiency terms to fund characteristics. The last section summarises and concludes.

2. Performance evaluation: methodology and results

In order to be able to compare investment results across funds, returns have to be corrected for risk. Since the seminal papers by Sharpe (1966), Treynor (1965) and Jensen (1968), several performance measures have been developed and applied empirically. Sharpe (1966) proposes to divide the
fund’s excess return by the standard deviation of its return, thus adjusting for total risk. However, as investors typically do not invest their entire wealth into one single fund, adjustment for priced, i.e. non-diversifiable, risk may be preferable. Using the CAPM as a benchmark, Treynor (1965) therefore adjusts excess return by the fund’s beta (see below). Similarly, Jensen (1968) proposes his alpha coefficient, i.e. the difference between the actual excess return and the expected excess return. For any fund \( p \) the Jensen alpha \( \alpha_p \) is given by:

\[
\alpha_p = R_p - E(R_p) = R_p - \beta_p E(R_m),
\]

where \( R_p \) is the actual average excess return of fund \( p \), \( \beta_p \) is its beta coefficient, \( R_m \) is the excess return on the market portfolio and \( E(\cdot) \) denotes the expectations operator. Excess returns are computed using a relevant risk-free rate. The second equality in (1) follows from assuming the CAPM to hold—but any other asset pricing model can be used. In a sample, \( \alpha_p \) can be estimated in a cross-sectional regression:

\[
R_p = \gamma_0 + \gamma_1 \hat{\beta}_p + \alpha_p,
\]

where the regression coefficients \( \gamma_0 \) and \( \gamma_1 \) should be zero and the average excess return on the market portfolio, respectively, if CAPM holds, and the hat on the beta coefficient indicates that it has to be estimated (the true beta is not observable).

Empirically, most papers have concentrated on the issue whether managed funds are able to outperform some relevant benchmark portfolios. Typically, the conclusion of these papers has been that, on average, funds did not significantly earn more than the passive benchmark performance on a risk-adjusted basis after deduction of expenses and commissions. Some authors even find that risk-adjusted fund returns are significantly negative. However, some individual funds were found to outperform the benchmarks significantly (see Ippolito (1993) for an overview of the literature). Of course, this does not necessarily contradict the market efficiency hypothesis since, purely due to sampling noise, it can be expected that some funds are found to 'outperform' in large samples.

Most studies have been carried out using data on domestic US mutual funds. But papers focusing on other markets reach similar conclusions. Detzler and Wiggins (1997) study the performance of 35 actively managed international funds using 111 monthly returns. Using a multi-index benchmark, they cannot reject the hypothesis that these funds exhibit no selectivity ability. Guy (1978) finds that British trusts (investing both domestically and internationally) do not significantly outperform the London Stock Exchange, nor randomly selected portfolios of UK and US stocks. Likewise, Cesari and Panetta (1998) investigate the performance of Italian mutual funds and find Jensen alphas that are usually not significant and in many cases even negative. Only when returns are gross of fees and costs do mutual funds show significant positive performance. Cai et al. (1997) analyse the performance of Japanese open-type stock mutual funds for the 1981–1992 period. Again, their results show considerable underperformance. This conclusion, and the one reached by Cesari and Panetta (1998), is independent from the benchmark model used. A similar conclusion for UK-based mutual funds is attained by Blake and Timmermann (1998) for the period 1972–1995 using the Micropal database.

3. Cross-sectional determinants

Although the literature on performance evaluation is abounding, relatively few papers try to relate performance to fund characteristics. Nevertheless, there is some theoretical and empirical evidence that these characteristics may be important. For instance, it may seem counter-intuitive that investors trust their capital to mutual funds, as these funds charge all kinds of fees and incur expenses. If markets are efficient and information on investment opportunities is widely available to all investors, fees should lower mutual fund returns below the return on investments managed by the investors themselves. In reality, information is not widely and freely available, so if the market for mutual funds were efficient, management fees should recover the costs of generating the necessary information (Grossman and Stiglitz, 1980). However, if agency problems exist, management
fees may exceed information costs and therefore lead to lower returns (Elton et al., 1993).

Ippolito (1989) finds evidence for US mutual funds consistent with the Grossman and Stiglitz (1980) hypothesis: funds charging larger fees also earn higher returns, and both effects off-set each other. However, Elton et al. (1993) report a number of data errors in Ippolito’s study. Correcting for these errors and measuring risk-adjusted returns using a multi-factor model, they reverse Ippolito’s conclusions and find that funds underperform their benchmarks. Droms and Walker (1994) find that US based international mutual funds cannot outperform their benchmarks. In contrast, in Droms and Walker (1996) the expense ratio of US domestic equity funds is found to be significantly related to fund performance: the higher the expense ratio, the higher the return. This would indicate that when security analysis is more intensively performed and/or more sophisticated strategies are pursued, the higher expenses are (more than) compensated by higher returns. However, the results in Malkiel (1995) indicate the conclusions reached by Droms and Walker (1996) are due to survivorship bias. Using a sample corrected for survivorship bias, Malkiel (1995) finds a significantly negative relation between expense ratios and raw returns for a sample of US equity funds over the period 1971–1991. Indro et al. (1999) confirm Malkiel’s results that US mutual funds underperform their benchmarks even gross of expenses. A weakly negative relation between management fees and risk-adjusted performance is also found for hedge funds (Ackermann et al., 1999).

A second characteristic that might be related to performance is size. Anecdotal evidence suggests that fund size may grow too large to be consistent with good management. In 1997, e.g., Fidelity Magellan, the world’s largest mutual fund, was closed to new investors, in order not to hamper investment management. Obviously, there are a number of reasons why big funds have a more difficult job to earn proper returns. Trading costs are sometimes higher for large-block transactions since the bid-ask spread increases dramatically with block size (Loeb, 1983). Indro et al. (1999) also argue that the attention large funds get from other market participants makes it more difficult to trade on information or to implement proprietary investment strategies. Large funds also need more managers, which may make the fund’s organisation more complex and more costly (Indro et al., 1999).

Related empirical evidence is reported by Dermine and Röller (1992). Their study is restricted to an investigation of economies of scale and scope of the French mutual fund industry. By estimating a cost function, they find that economies of scale and scope exist for smaller firms and across all fund categories. Their result indicates that the mutual funds’ size—or rather their sponsor firm’s size—is potentially an important characteristic when fund performance is to be explained cross-sectionally. This is confirmed by Chen et al. (1992) who find a significantly positive relation between the performance (measured by Jensen’s alpha) of US domestic stock funds and their size. More detailed results in Indro et al. (1999) point to the existence of an increasing relation between fund size and return except for the funds in the largest size decile, where returns drop. A size effect is also reported in Dahlquist et al. (1999) for Swedish mutual funds. Large equity funds seem to perform worse than smaller funds. However, the reverse holds for bond funds, which are relatively small in Sweden. Elton et al. (1996) find that the relationship between size and performance may be affected by survivorship bias. When this possible bias is ignored, there is no difference in performance between large and small funds. However, when survivorship bias is taken into account, smaller funds tend to underperform larger funds. The bias is attributed to the fact that smaller firms face a higher probability to disappear.

A third characteristic is suggested by the literature on performance persistence. Hendricks et al. (1993) and Goetzmann and Ibbotson (1994) both find that winning funds (“winners”) over a reference period are more likely to be in the better performing group in the subsequent period than in the worse performing group (“losers”). Therefore, a simple ranking of funds based on historical performance may contain information for investors about future performance. However, up to now it is not clear whether or not performance
persistence is an artefact of the data. Malkiel (1995) finds that the presence of persistence is dependent on the sample period; persistence is found in the 1970s, but not in the late 1980s. Moreover, Brown et al. (1992) argue that performance persistence may be due to survivorship bias in the sample of mutual funds. If funds that disappear during the sample period are not taken into account in the persistence analysis, apparent persistence may be found, even if in reality performance is independent across time. On the other hand, Hendricks et al. (1993) and Brown and Goetzmann (1995) still find performance persistence in samples of US equity mutual funds after controlling for survivorship bias as much as possible. It should be stressed that this persistence may be due to risk factors that are not accurately taken into account. Some corroboration of this view is found in Sauer (1997) who does find performance persistence for US equity funds, at least when no distinction is made according to the funds' investment objectives. When the sample is partitioned by investment style, the persistence disappears. However, using conditional alphas and controlling for survivorship bias as much as possible, Christopherson et al. (1998) find significant performance persistence for institutional equity managers. This persistence is mostly due to losers who tend to stay losers in the next period. This is also reported in Grinblatt and Titman (1992) and in Carhart (1997) who, in a comprehensive sample of US equity funds, are only able to explain the risk-adjusted return persistence of losers. As for international mutual funds, Detzler and Wiggins (1997) find no significant performance persistence, whereas Blake and Timmermann (1998) report some performance persistence for UK mutual funds. Given these mixed results we use historical measures of performance in our study to investigate the link with efficiency measures.

Finally, fund age (period since inception) may also be a determinant of fund performance to the extent that economies of experience are important. The empirical literature has devoted little attention to this issue. In their hedge fund study, Ackermann et al. (1999) do not find any significant relation between age and risk-adjusted performance measured by the Sharpe ratio. Contrary, Blake and Timmermann (1998) investigate the abnormal performance of UK mutual funds in the periods just before the funds' death or shortly after their birth. During the final year of existence the fund performs 3.3% worse than the average fund. A small but short-lived positive outperformance of 0.8% is found for the first year of existence. Evidently, such a positive relation between age and performance during the first years of existence of the fund may indicate an experience effect. On the other hand, it may also indicate survivorship bias as older firms are probably only included in the database if their performance was sufficiently high.

4. Data description

This paper presents new evidence on these issues by using a European sample of equity mutual funds and by applying a different kind of methodology to compute mutual fund efficiency. Since the liberalisation of European stock markets and the deregulation of capital account transactions, cross-border stock investments have increased substantially. The introduction of the euro has accelerated this trend and will continue to do so in the future (McCauley and White, 1997; Giavazzi et al., 2000). The existing European literature has predominantly studied either domestically oriented equity funds or funds that invest internationally but are marketed only in the sponsor's home country. As the European mutual fund business has become more competitive, a broader focus is needed. Our sample contains all mutual funds that were classified by Micropal as “European Equity Funds” and were included in the Micropal Databases Off-shore Territories, France, and Benelux on the 31st July 1998. Classification is done based on the investment objectives stated by the funds. We compute monthly returns by dividing the fund's end-of-month net asset value per share (NAV) by the previous month's NAV and subtracting one. The NAV data are adjusted to include dividends or other capital distributions. To make all returns comparable, all NAV are converted in ECU. Because of the requirement that at least 24 monthly returns should be available, we
retain 179 European equity funds from the 345 available series. 4

In Table 1 cross-sectional summary statistics for the 179 mutual funds are shown. All statistics are computed using the last 36 monthly data (August 1995–July 1998), except for the funds with fewer returns. 5 Average returns vary between 0.12% and 3.18% per month, with an average (median) of 1.57% (1.60%). This compares to an average return of 2.10% for the European stock market index constructed by Datastream over the corresponding 36 months. As a first indication of performance, we compute Jensen’s alpha for each fund. Alpha is obtained as the constant from a time-series regression of the fund’s excess return upon a constant and the excess return on Datastream’s Pan-European index (\(R_{mt}\)). 6 Excess returns are computed using the 1-month ECU LIBOR as the risk-free rate.

The funds’ alpha coefficients are found in the second data row of Table 1. In addition, the table shows the summary statistics for the funds’ beta coefficients (third row) as well as the regression \(R\)-squared coefficients (last row). Although not reported, the regression equations are generally highly significant, with high \(R\)-squared (median 89%) and significant beta coefficients; only one beta is not significant at the 5% level. The betas centre around one, indicating the relevance of our benchmark. For the sample as a whole, the average alpha coefficient, being close to zero, does not indicate significant outperformance. The average alpha is in line with the results in Droms and Walker (1994) who find an average alpha of \(0.44\%\) per annum, which is equivalent to \(0.037\%\) per month. In our sample 60.9% of all alphas are negative. Based on a one-sided univariate \(t\)-test, 5 out of 179 alpha coefficients are significantly positive at the 5% level, whereas 28 are significantly negative. This predominance of negative alphas is even more remarkable given the fact that our sample is subject to survivorship bias and returns therefore contain an upward bias. In the US, survivorship bias can be estimated at 0.9–1% per annum (Elton et al., 1996; Malkiel, 1995), whereas Blake and Timmermann (1998) find a bias of 0.78% per annum for UK-based international growth equity funds. Dahlquist et al. (1999) estimate the survivorship bias for Swedish equity funds between 0.6% and 0.7% per annum.

Moreover, the statistics reported in Table 1 may depend upon the benchmark used in the regression, as well as the estimation period. Indeed, many papers indicate that the use of other benchmarks or multiple benchmarks can dramatically alter the

Table 1

Cross-sectional summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average return</td>
<td>1.57%</td>
<td>1.60%</td>
<td>0.38%</td>
<td>0.12%</td>
<td>3.18%</td>
</tr>
<tr>
<td>Alpha</td>
<td>−0.04%</td>
<td>−0.10%</td>
<td>0.44%</td>
<td>−1.42%</td>
<td>1.88%</td>
</tr>
<tr>
<td>Beta</td>
<td>0.94</td>
<td>0.99</td>
<td>0.16</td>
<td>0.39</td>
<td>1.24</td>
</tr>
<tr>
<td>(R)-squared</td>
<td>0.82</td>
<td>0.89</td>
<td>0.18</td>
<td>0.03</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Returns and alphas are in percentages per month. Average return is the average return over the last 36 months (August 1995–July 1998), unless fewer observations were available. At least 24 monthly returns were used. Alpha (beta) is the intercept (slope) coefficient from a regression of the fund’s return in excess of 1 month ECU LIBOR on the excess return of the Datastream European stock market return index. \(R\)-squared is the determination coefficient of this regression.

---

4 This sample is subject to some biases. First, the sample does not cover the entire market, but only the funds commercialised in the geographic areas mentioned above. Moreover, it is not clear whether the Micropal universe covers even this area entirely. Second, the sample most likely suffers from survivorship bias. This can be inferred from the fact that not a single fund in our sample ceased to exist. It is not sure how this affects our analysis. If anything, our efficiency measures will be biased upwards, as the most inefficient funds are not contained in our sample.

5 Given our sample criteria, at least 24 observations were available to compute the summary statistics in Table 1.

6 When the CAPM holds and \(R_m\) is a good proxy for the market portfolio return, this procedure is equivalent to Eq. (2).
performance results.\(^7\) We experimented with different benchmarks as well as with different estimation periods and obtained qualitatively similar results.\(^8\)

5. Stochastic frontier methodology

Instead of simply calculating excess returns, our methodology is designed to capture the deviation of the funds’ performance from what an efficient (best-practice) portfolio can achieve. Since most performance studies conclude that no positive abnormal returns can systematically be earned by mutual funds, we treat this finding as a maintained assumption. This is in accordance with the EMH, which predicts that the expected abnormal return should be zero, or even negative if the fund incurs excessive transaction and management costs. Yet, in any sample it is likely that significantly positive abnormal returns (or alphas) are found for some funds, purely due to sampling noise. To increase statistical power, we augment Eq. (2)\(^9\) with a composed error, which consists of a symmetric disturbance capturing measurement error (\(\nu_p\)) and a non-negative disturbance term, modelling the level of efficiency (\(\zeta_p\)):\(^10\)

\[
R_p = \gamma_0 + \gamma_1 \tilde{\beta}_p - \zeta_p + \nu_p.
\] (3)

Both error terms are assumed to be independent from each other and across funds. The non-negativity constraint on \(\zeta\) follows from the (empirically documented) assumption that a managed fund cannot systematically outperform the benchmark portfolio. The exact estimation procedure for \(\tilde{\beta}_p\) will be discussed in the next section.\(^11\)

In order to avoid the requirement of having to impose an arbitrary distribution for the non-negative disturbance, i.e. the efficiency measure, we propose to use various distributions and pool them. Mixing different distributions for each individual efficiency within the sampling model would make the analysis intractable. Instead, we use the Bayesian approach because the pooling approach is quite natural in a Bayesian framework. Here, we deal with simple tractable models and mix them in the final stage. Therefore, we adopt the Bayesian frontier approach as developed by van den Broeck et al. (1994). We assume that \(\nu_p\) is normally distributed with mean zero and variance \(\sigma^2\), \(\zeta_p\) is gamma distributed with shape parameter \(j\) and unknown scale parameter \(\lambda\), leading to different statistical distributions. Taking into account that a gamma distribution with a high value for the shape parameter \((j)\) makes the shapes of the densities of \(\nu_p\) and \(\zeta_p\) hardly distinguishable, the adopted Bayesian frontier approach restricts itself to three gamma distributions with small integer values for the shape parameter \((j = 1, 2, 3)\) as competing hypotheses. The underlying idea is to create a very flexible distribution for the efficiency disturbance term by a pooling process, which is still tractable and not to look for the ultimate correct distribution for the efficiency term. These three distributions are referred to as Erlang distributions.

By attaching prior probabilities to the parameters of the Erlang models, which do not affect

---

\(^7\) See e.g., the difference between the results in Ippolito (1989) and those in Elton et al. (1993) The latter use a three-index benchmark, which accounts for a small stock and a bond effect. The positive performances found by Ippolito (who uses only the S&P 500 index as benchmark) disappear or even turn negative in Elton et al. (1993). For a more extensive discussion of this issue, see Grinblatt and Titman (1995).

\(^8\) We also used the MSCI Europe 15, MSCI Europe ex UK and Datastream Europe ex UK indices as benchmarks. Correlations between betas using these benchmarks were nearly one. Estimation periods of 24, 48 and 60 months were also investigated. Again very high correlations across estimation periods were found. Results are available upon request.\(^9\)

\(^9\) Eq. (2) is based on the static CAPM of Sharpe (1964) and Lintner (1965). It is evident that other equilibrium asset pricing models (see e.g., Fama and French (1992)), can easily be integrated in our approach.

\(^10\) It might be argued that our distributional assumption on \(\zeta\) does not confirm with our Eq. (1), where \(z\) is seen to have a zero mean. However, this is only in the case when all funds are perfectly efficient, an assumption which is clearly not supported by previous research. We thank a referee for pointing out this potential point of confusion.

\(^11\) Note that we do not take into account the estimation error attached to our \(\tilde{\beta}_p\) estimates in our stochastic frontier methodology. As we obtained relatively robust estimates for this coefficient (see Section 6), we do not think this will invalidate our results.
the salient features of the different models, the Bayesian approach leads to their posterior probabilities. In our case we use the following prior densities

\[ p(\gamma, \sigma^2, \lambda) = c\sigma^{-2} \lambda^{-1} \]  

(4)

These are diffuse priors, because we know only the range of the parameters (i.e. Jeffreys type prior). 12 This is true for each Erlang model, because no extra information is available. This means that the diffuseness for all the parameters, which are common to all three models, remains.

Following this Bayesian procedure, we do not need to choose among the competing models. We can average out specification uncertainty by mixing individual posterior densities \( P(M_j|y,X) \) of a quantity of interest (\( \psi \)) into its overall posterior density \( p \):

\[ p(\psi|y,x) = \sum_j P(M_j|y,X)p_j(\psi|y,X) \]  

(5)

Instead of artificially imposing one specific distribution for the efficiency term, we mix over simple models differing in the distribution of all the inefficiency terms.

Another reason to proceed this way is of a more philosophical nature. The tradition of modelling is largely focussed on searching a “true” model, and by consequence rejecting the “false” models. By pooling different models, we treat them as joint providers of valuable, but probably partial, information, dismissing the very idea of the existence of a single “true” model (Winkler, 1989). The pooling of the various distributions creates a new distribution, which generates new estimates of the relevant parameters. The latter are the result of the aggregation of the weighted information contained in the original distributions, which reduces uncertainty and increases statistical accuracy. The individual efficiency (\( r_p \)) measure is expressed as \( r_p = \exp(-\xi_p) \), whereas the average efficiency \( r_f \) is the cross-sectional average over all \( r_p \).

For benchmark reasons we also estimate a baseline model \( M_0 \) which is the simple case without composed error, i.e. all \( \xi_p \) are zero and thus all funds are considered to be on the frontier. By doing so we can evaluate the improvements of our composed error models \( M_j (j = 1, 2, 3) \) against the baseline model \( M_0 \) e.g., the neutral or non-neutral character of the shift of the frontier.

6. Stochastic frontier implementation and results

Our investigation of Eq. (3) starts with the static CAPM of Sharpe (1964) and Lintner (1965). The static CAPM implies that returns on mutual funds are proportional to their systematic risk (\( \beta_p \)). Although our results are not dependent on the chosen benchmark (see Section 4), it is possible that our analysis is biased because we include funds whose stated investment objectives do not coincide with the actual ones. Indeed, as Table 1 shows, some of the fund returns are hardly related to the benchmark return. We therefore discard the three funds with R-squared coefficients lower than 30%. 13 In addition, in order to avoid well-known biases in the estimation of the Security Market Line (e.g., Elton and Gruber, 1995), we estimate beta coefficients by regressing fund excess returns on \( R_{mt} \) over a previous estimation period. The beta for a fund is then the estimated slope coefficient. More specifically, we use the previous 24–36 months (depending on data availability) to compute betas. These betas are then related to the funds’ average returns over the following 36 months. 14 This procedure produces a reduction in the number of observations since only 134 funds have sufficient data to estimate relevant beta coefficients. All the results below refer to this sample. To ensure that our empirical results do not

12 In case of diffuse priors, i.e. when only little prior information is available such as the range of the parameters, Jeffreys’ and Haldane’s rules are adopted. These are (Jeffreys, 1967)

\[
\begin{align*}
-\infty < \theta < +\infty & \quad \text{constant (Jeffreys’ rule)} \\
0 \leq \theta < +\infty & \quad \theta^{-1} (\text{Jeffreys’ rule}) \\
0 \leq \theta \leq 1 & \quad (\theta(1 - \theta))^{-1} (\text{Haldane’s rule})
\end{align*}
\]

For the connection between these rules and the maximum likelihood estimators see Broeckx et al. (1974).

13 Higher cut-off values were also used with similar results.

14 Beta coefficients are relatively stable across the two periods: the correlation coefficient between the two sets of betas is 0.65.
rely too heavily on the benchmark model (CAPM), we also estimated a multi-factor model in the spirit of Fama and French (1993). As we found nearly identical results we refrain from reporting them.

In order to calculate the parameters of the three Erlang models \( M_j, \; j = 1, 2, 3 \) a computer program BSFM, based on a Monte Carlo technique, has been used (Arickx et al., 1997). The starting values for the parameters were obtained by estimating the baseline model \( M_0 \) (without composed error). The results are reported in the first column of Table 2.

After the calculations were performed, including the calculation of the individual funds’ efficiencies, the results are pooled by applying equal weights to each of the models, i.e. \( 1/3 \). Integrating out uncertainty gives the weight for each separate model. The pooled results are the sum of the weighted parameters of each Erlang model and are also reported in Table 2. The estimated densities for the three Erlang models as well as the pooled density are shown in Fig. 1.

Table 2 shows the posterior moments of the model parameters with diffuse prior. All estimations produce statistically significant coefficients.

Comparing the results of \( M_0 \) with the Erlang model, the frontier appears to be a non-neutral shift of the average function, that is the intercept \( \gamma_0 \) shows an upward shift and the slopes \( \gamma_1 \) have a lower value. However, the statistical significance of the latter can be discussed. Disregarding the non-neutral character of the shift of the frontier, the shift itself is situated between 31.5% and 57.0% of the average relationship (i.e. without composed error). For the pooled results the upward shift amounts to 44.1%, which is quite considerable.

The standard error of the normally distributed statistical noise is denoted \( \sigma \) and \( \lambda \) is a parameter for the efficiency measure (van den Broeck et al., 1994). \( TV_f \) represents the posterior out-of-sample error variance and \( VR_f \) indicates the relative importance of the statistical error \( \mu_p \) in the total error variance. The latter suggests that the statistical noise component is dominant and needs to be filtered out to get a less contaminated efficiency measure. This is precisely what the stochastic frontier approach achieves.

In the bottom panel of Table 2 the log-likelihood for each model \( M_j \) (Log \( l_j \)) are provided. Giving each Erlang model the same prior weight, i.e. \( P(M_j) = \frac{1}{3} \), and integrating out model uncertainty, the posterior probabilities for Erlang 1, 2 and 3 are respectively 35.8%, 32.4% and 31.8%. This means that Erlang 1 (i.e. the exponential distribution) is slightly more important but is not

---

15 Interested readers may refer to a previous working paper version, which is available upon request.
dominating, because the other Erlang distributions remain present in the pooling process. Apparently the data are dictating that the true distribution is a complex one indicating that many mutual funds are severely underperforming. These results demonstrate the power of the pooling approach vis-à-vis the selection of one specific model out of several possible alternatives.

Besides the pooling procedure, another feature of the Bayesian stochastic frontier approach is the direct way of estimating the individual efficiencies. In Table 3 we represent the posterior out-of-sample (average) efficiency \( r_j \) for each Erlang model as well as the pooled efficiencies and the lowest and highest estimates of the individual fund efficiencies. The individual efficiencies exhibit large differences across the sample of mutual funds. The fund with the lowest efficiency is located approximately at a level of 50% below the efficient frontier (see second data row of Table 3). The funds with the highest level of efficiency are positioned near the frontier. These results clearly show the substantial variation in fund efficiencies, hence the need to investigate the determinants of fund efficiency.

7. Determinants of efficiency

In this section, we relate the relative efficiency scores of the mutual funds to some of their characteristics. We focus on the relationship between efficiency and size, historical performance measures, and fund age.\(^{16}\) We also experimented with

---

\(^{16}\) Unfortunately, one of most interesting attributes, i.e. the expense rate, is not available for our sample of funds.
the fund’s nationality and its sponsor, but did not find any significant results.

7.1. Size

First, to alleviate the reverse causation problem, we only use the funds for which the size was reported for December 1995 (or if missing, the size at the end of any earlier month) and relate these size figures to the funds’ efficiency scores. The drawback of this approach is a loss of information, since the size figure for 1995 is only available for 81 funds. Alternatively, we relate fund efficiency to fund size in later years (December 1996 and December 1997), thus increasing the sample to respectively 117 and 123 funds. However, this approach also has the drawback that, if a positive relation is found between size and efficiency, it is not clear in which direction the causation runs. For instance, Sirri and Tufano (1998) find that exceptionally good performance leads to large inflows of new money (the reverse does not occur for extremely poor performance).

Table 4 reports the Pearson correlation coefficients as well as Spearman rank correlations between the relevant variables. To reduce skewness of the size variables, they are transformed by taking natural logarithms.\(^{17}\)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>–</td>
<td>0.172</td>
<td>0.252</td>
<td>0.325</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.124)</td>
<td>(0.006)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Ln(size 1995)</td>
<td>0.215</td>
<td>–</td>
<td>0.668</td>
<td>0.690</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Ln(size 1996)</td>
<td>0.288</td>
<td>0.867</td>
<td>–</td>
<td>0.907</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.000)</td>
<td></td>
<td>(0.000)</td>
</tr>
<tr>
<td>Ln(size 1997)</td>
<td>0.373</td>
<td>0.863</td>
<td>0.919</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td></td>
</tr>
</tbody>
</table>

Entries above the main diagonal denote Pearson correlation coefficients, below the main diagonal Spearman rank correlation coefficients. Numbers between parentheses are marginal (two-sided) significance levels. Correlations with the 1995 size variable are based on 81 observations, those with the 1996 size variable on 117 observations, and those with the 1997 size variable on 123 observations.

Based on these correlations, a positive relationship is found between efficiency and size. This is most obvious for the 1997 figures, but as argued above, it is not clear whether the causation runs from size to efficiency. The relation is weaker for the 1995 figures, which may be due to outliers and the smaller number of observations. Indeed, the rank correlation is higher than the Pearson correlation coefficient and marginally significant at the conventional 5% level.

To investigate potential non-linear relations between size and efficiency, we divide the data into quintiles, based on breakpoints using the size variables. For each quintile we compute mean and median efficiency. The results can be found in Fig. 2. Panel (a) graphs average and median efficiencies per 1997 size quintile, whereas the comparable figures for the 1995 size quintile are reported in panel (b). The 1997 results show a low and declining level for the first two quintiles. For larger sizes efficiency increases up to the last quintile (largest funds). This finding holds both for average efficiency and median efficiency (although for the latter there is a drop in the fourth quintile). The average efficiencies are significantly different from each other according to a standard F-test (marginal significance 0.2%). Likewise, a Kruskal-Wallis test indicates that the median values are also significantly different across quintiles (marginal significance 0.0%).\(^{18}\)

The picture is slightly different when we use the 1995 size breakpoints. There is no longer a decrease of efficiency in the second quintile, and efficiency generally increases monotonically. However, differences in mean nor in median efficiencies are statistically different across quintiles. This is not consistent with the results of Indro et al. (1999) who find declining risk-adjusted returns for the largest size decile.

\(^{17}\) Of course, efficiency terms are also highly skewed. We therefore repeated the analysis using logit-transformed efficiency terms. All results reported for the raw efficiency measures also hold for the transformed terms.

\(^{18}\) Results are similar when breakpoints for the 1996 size variable are used.
Overall, the evidence suggests that efficiency is positively related to fund size, which may point to the existence of economies of scale in the mutual fund industry. It can be argued that the weak relation found between the 1995 size figures and subsequent efficiency scores may be an indication that causation runs in the other direction: good performance leads to more capital inflow.

7.2. Performance persistence

Given the ample empirical evidence that historical performance contains some indication about future performance, we relate efficiency to several measures of historical performance. We would have liked to use a historical estimate of our efficiency estimate, but because of lack of sufficient historical data, this was not possible. We therefore use more traditional performance

---

19 We also regressed efficiency on size. Both for 1997 and 1996 we found a significant positive coefficient for size, but not for the 1995 data. However, the $R^2$-squared coefficients are relatively low (maximum of 11%). Introducing a quadratic size variable as an additional regressor did not improve the fit and the coefficient on the quadratic term was never significant.
measures. First, we use the historical alpha coefficient. This coefficient is obtained as a by-product of the time series regression in Section 4. However, since it has been demonstrated that performance persistence is a short-run phenomenon, alpha coefficients that relate to a 36 month-period may be too stale. Because estimating alpha coefficients over shorter horizons is likely to lead to noisy estimates, we prefer to use average historical return instead. More specifically, we use one-year average returns and, if sufficient data are available, also two-year and three-year average returns.

In Table 5 correlations between the return momentum measures and efficiency are shown. As expected, all performance measures are significantly positively related. The only risk-adjusted measure, the alpha coefficient, is only weakly related to efficiency; only the rank correlation is significantly different from zero at conventional levels. This may be due to the short-run behavior of performance persistence. Indeed, the shorter the period over which returns are averaged, the closer the relationship with efficiency becomes. Returns computed over 12 months are significantly related to efficiency.

Table 5
Correlations with historical performance measures

<table>
<thead>
<tr>
<th></th>
<th>Efficiency</th>
<th>Return previous 12 months</th>
<th>Return previous 24 months</th>
<th>Return previous 36 months</th>
<th>Historical alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>–</td>
<td>0.182 (0.035)</td>
<td>0.157 (0.070)</td>
<td>0.127 (0.178)</td>
<td>0.032 (0.711)</td>
</tr>
<tr>
<td>Return previous 12 months</td>
<td>0.231 (0.007)</td>
<td>–</td>
<td>0.643 (0.000)</td>
<td>0.246 (0.008)</td>
<td>0.391 (0.000)</td>
</tr>
<tr>
<td>Return previous 24 months</td>
<td>0.201 (0.020)</td>
<td>0.583 (0.000)</td>
<td>–</td>
<td>0.698 (0.000)</td>
<td>0.748 (0.000)</td>
</tr>
<tr>
<td>Return previous 36 months</td>
<td>0.166 (0.077)</td>
<td>0.259 (0.005)</td>
<td>0.642 (0.000)</td>
<td>–</td>
<td>0.926 (0.000)</td>
</tr>
<tr>
<td>Historical alpha</td>
<td>0.181 (0.036)</td>
<td>0.444 (0.000)</td>
<td>0.756 (0.000)</td>
<td>0.865 (0.000)</td>
<td>–</td>
</tr>
</tbody>
</table>

Entries above the main diagonal denote Pearson correlation coefficients, below the main diagonal Spearman rank correlation coefficients. Numbers between parentheses are marginal (two-sided) significance levels. All return figures are average returns computed over the period indicated. All correlations are based on observations for all 134 funds, except those with the 36-month average return, which is based on 114 funds.

Fig. 3. Mean and median efficiency per 12 month return quintile. The graph shows mean (left bars) and median (right bars) efficiency per return quintile. The funds’ returns over the previous 12 months were used to determine the quintile breakpoints.
This positive relation can also be seen in Fig. 3, where average and median efficiency per 12 month return quintile is depicted. However, an ANOVA indicates no significant average differences across quintiles (the medians are different at a marginal significance level of 8%). This is hardly surprising when we look at Fig. 3, as only the first quintile has a substantially lower average efficiency than the others. This is consistent with the evidence that performance persistence is to a large extent due to the “losers”: poor performers stay poor performers. A Mann-Whitney test indicates that median efficiency is significantly lower in the first quintile than in the others (marginal significance 3.3%).

7.3. Fund age

The Micropal database also reports the starting date for each fund, which allows us to compute their age. The Pearson correlation coefficient between age and efficiency is 0.124, which is not significant at the 5% level. Neither does the rank correlation coefficient (0.122) point to a significant relation. Moreover, the average and median efficiency per age quintile (Fig. 4) does not disclose an obvious association between age and efficiency. Neither the \( F \)-statistic from an ANOVA analysis nor the Kruskal-Wallis test reveal significant differences in average resp. median efficiency across age quintiles. This is consistent with the hedge fund study by Ackermann et al. (1999). The positive outperformance of mutual funds in the first year of their existence, which is documented in Blake and Timmermann (1998), could not be tested as our procedures to compute fund efficiency require more than 12 months of historical data.

8. Conclusions

In this paper we acknowledge the results from three decades of mutual fund performance studies that mutual funds are not able to outperform passive benchmarks. Nevertheless, cross-sectional differences in performance may exist and identification of such differences may be important to investors and the mutual fund industry. The purpose of this paper is to identify readily available ex ante fund statistics that can be related to future performance of European equity funds over the period 1995–1998. This task is hampered by the
fact that ex post, purely due to chance, some funds are found to outperform the benchmark portfolios. In this paper we therefore decompose the deviation from their expected return into a noise component and an efficiency score, which would be 100% if the fund exhibited no underperformance. The decomposition is based on the Bayesian frontier approach as developed by van den Broeck et al. (1994).

We find that for a sample of European equity funds efficiency is positively related to fund size. Large funds outperform small funds, which may indicate the presence of scale economies in the European equity mutual fund industry. However, as we rely on an ex post size figure, the relationship may also be due to a flow effect: investors may base their investment decision on historical performance. This would also result in a positive relationship between performance and ex post size. Nevertheless, the positive relationship found for the ex ante size figures provides evidence for the former explanation, but due to their poor quality the statistical evidence is weaker.

In addition to a relation with size, we also find that performance is related to historical performance, which is in line with the performance persistence literature. This effect is almost entirely due to the poorly performing group of funds. No relationship between efficiency and historical return is found in the top 80% of funds. Finally, we fail to find a link between fund age and performance.

This paper is certainly prone to improvement. First of all, the data can be supplemented from other sources. Especially expense variables need to be included more comprehensively. Also the methodology may be improved upon. The efficiency measures rely upon the cross-sectional relation between risk and expected return as implied by a given asset pricing model. In this paper, we used the static CAPM—and also a two factor model including a High Minus Low portfolio in the spirit of Fama and French (1993) although these results were not explicitly reported. To the extent that other pervasive factors are present in the European stock market, other efficiency estimates may be obtained. As the search for a comprehensive European asset pricing model is beyond the scope of this paper we leave this for future research.

Acknowledgements

We express our gratitude to Elvira Haezendonck for her assistance with the calculations. Comments of the participants of the 6th European Workshop on Efficiency and Productivity Analysis in Copenhagen and the North American Productivity Workshop at Union College in Schenectady are gratefully acknowledged. We also thank three anonymous referees for their constructive remarks and suggestions. This research was initiated when the first author was affiliated to Erasmus University Rotterdam (the Netherlands).

References


